Implicit Function Theorems for Lipschitz Functions

by

Sunny Sood

MA4K9 Dissertation

Submitted to The University of Warwick

Mathematics Institute

April, 2020

Contents

1 Introduction

The function $f : U \to \mathbb{R}^m$, where $U \subseteq \mathbb{R}^n$ open is called *Lipschitz* if there exists $K > 0$ such that for any $x, y \in U$,

$$\|f(x) - f(y)\| \leq K \|x - y\| .$$

We call K the *Lipschitz constant* of f in U. An alternative notation for the Lipschitz constant of f is $\mathrm{Lip}(f)$. An example of a Lipschitz function is the norm function from $\mathbb{R}^n$ to $\mathbb{R}^n$, with Lipschitz constant 1.

It is not necessarily true that Lipschitz functions are everywhere differentiable in the classical sense. For example, the norm function is not differentiable at the origin. Remarkably however, Lipschitz functions are differentiable almost everywhere. This is the statement of *Rademacher's theorem*, for which we have provided a proof in this report. Using Rademacher's theorem, it is possible to well define the notion of the *generalised derivative* of a Lipschitz function. This notion of derivative is due to F.H.Clarke and first appears in his paper [1]. Let us define this derivative:

Definition 1.1. *Let $f : U \to \mathbb{R}^m$ be Lipschitz, $U \subseteq \mathbb{R}^n$ open and $x_0 \in U$. Then, we define the generalised derivative at x_0, denoted $\partial f(x_0)$, as*

$$\partial f(x_0) := convH\left(\left\{\lim_{m\to\infty} d_{x_n} f \mid x_n \to x_0\right\}\right)$$

where we consider all possible x_n for which $x_n \to x_0$, the classical derivative $d_{x_n} f$ is defined and $\lim_{m\to\infty} d_{x_n} f$ exists. The notation convH denotes taking the convex hull in the space of $m \times n$ real matrices.

We want to show that the generalised derivative is well defined. By Rademacher, f is differentiable almost everywhere in U. In particular, taking balls centred at x_0 of decreasing radius, we deduce that there exists a sequence $(x_n)_{n=1}^{\infty} \subset U$ such that $x_n \to x_0$ and $d_{x_n} f$ exists in the classical sense. Furthermore, as f is Lipschitz in U, it follows that there exists $K > 0$ such that for any $x, y \in U$,

$$\|f(x) - f(y)\| \leq K \|x - y\| .$$

In particular, for $x \neq y$, we have

$$\frac{\|f(x) - f(y)\|}{\|x - y\|} \leq K.$$

Therefore, where the partial derivatives of f exist, their norms are bounded above by K. Thus, as the set of $m \times n$ real matrices with bounded supremum norm is a compact set and

$$\{d_{x_n} f \mid x_n \to x_0\}$$

is a sequence in this set, by sequential compactness, this sequence must contain a convergent subsequence. Hence, $\partial f(x_0) \neq \emptyset$. Furthermore, the set

$$\left\{ \lim_{m\to\infty} d_{x_n} f \mid x_n \to x_0 \right\}$$

is bounded because f is Lipschitz at x_0.

Now, Clarke claims that the above set is "obviously" closed. We think that he had a diagonal argument for sequences in mind. However, one needs to be careful when using such arguments, as the list of sequences

$$\begin{gathered} (1, 0, 0, 0, \dots) \\ (0, 1, 0, 0, \dots) \\ (0, 0, 1, 0, \dots) \end{gathered}$$

demonstrates. We will continue assuming that he is correct (otherwise, one may replace the set $\{\lim_{m\to\infty} d_{x_n} f \mid x_n \to x_0\}$ by its closure).

With this, it follows that $\{\lim_{m\to\infty} d_{x_n} f \mid x_n \to x_0\}$ is bounded and closed and therefore compact. Thus, $\partial f(x_0)$ is a convex and compact, as the convex hull of a compact set in finite dimensional space is compact (see [6], Theorem 17.2).

Before we explain our motivations for this report, let us study an example to see how we may compute the generalised derivative.

Let us consider the function

$$\begin{aligned} f &: \mathbb{R}^2 \to \mathbb{R}^2 \\ f(x, y) &:= (|x| + y, 2x + |y|)\,. \end{aligned}$$

Note, the derivative of f is defined everywhere in the classical sense except at the origin. We seek to compute $\partial f((0, 0))$.

Suppose $(x_n, y_n) \to (0, 0)$ where $x_n < 0$ and $y_n > 0$. Then, $f(x_n, y_n) = (-x_n + y_n, 2x_n + y_n)$. Therefore, in this case, we have

$$d_{(x_n, y_n)} f = \begin{pmatrix} -1 & 1 \\ 2 & 1 \end{pmatrix}.$$

Considering all other possible cases similarly, we conclude that

$$\begin{aligned} \partial f((0, 0)) &= \mathrm{convH}\left(\left\{ \begin{pmatrix} -1 & 1 \\ 2 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 1 \\ 2 & 1 \end{pmatrix}, \begin{pmatrix} -1 & 1 \\ 2 & -1 \end{pmatrix}, \begin{pmatrix} 1 & 1 \\ 2 & -1 \end{pmatrix} \right\} \right) \\ &= \left\{ \begin{pmatrix} s & 1 \\ 2 & t \end{pmatrix} : -1 \leq s, t \leq 1 \right\}. \end{aligned}$$

The calculus of generalised derivatives is studied extensively in [2]. Of interest to us is an

Inverse Function Theorem of Lipschitz functions, also stated and proved in [2][pp 253-255]:

Theorem 1.2. *Let $f : \mathbb{R}^n \to \mathbb{R}^n$ be Lipschitz, $x_0 \in \mathbb{R}^n$. If every $A \in \partial f(x_0)$ is invertible, then f is locally invertible with Lipschitz inverse. That is to say, there exists open neighbourhoods U of x_0, V of $f(x_0)$ and Lipschitz $g : V \to \mathbb{R}^n$ such that $g \circ f(u) = u$ for all $u \in U$ and $f \circ g(v) = v$ for all $v \in V$.*

We will use this Inverse Function Theorem to prove the following Implicit Function Theorem:

Theorem 1.3. *Let $U \subseteq \mathbb{R}^{n+k}$ be open; $(x, y) = (x^1, \cdots, x^n, y^1, \cdots, y^k)$; let $f : U \to \mathbb{R}^k$ be Lipschitz at (a, b) in U and define $c := f(a, b)$. If there exists a $n \times (n + k)$ matrix B such that for every $A \in \partial f((a, b))$, $\begin{pmatrix} A \\ B \end{pmatrix}$ is invertible, then $f^{-1}(c)$ is locally a Lipschitz submanifold of dim n.*

We remark that this Implicit Function Theorem will be one of three Implicit function theorems for Lipschitz functions that appear in this report. It is natural to ask if and how they are related to each other. We have started to explore this question in this report, but we believe our investigations are not complete.

We want to highlight the hypothesis of this theorem: **If there exists a** $n \times (n + k)$ **matrix** B **such that for every** $A \in \partial f((a, b))$**,** $\begin{pmatrix} A \\ B \end{pmatrix}$ **is invertible**, then $f^{-1}(c)$ is locally a Lipschitz submanifold of dim n. In practise, it is often easier to show that for every $A \in \partial f((a, b))$, there exists a $n \times (n + k)$ matrix B_A such that $\begin{pmatrix} A \\ B_A \end{pmatrix}$ is invertible. This is equivalent to saying that every $A \in \partial f((a, b))$ has maximal rank. Thus, it is natural to consider the following problem:

Problem 1.4. *Let $U \subseteq \mathbb{R}^n$ be open; $n > p$; $f : U \to \mathbb{R}^p$ Lipschitz and $x_0 \in U$. Then, if for every $A \in \partial f(x_0)$, there exists a $p \times n$ matrix B_A such that $\begin{pmatrix} A \\ B_A \end{pmatrix}$ is invertible, is it true that there exists a $p \times n$ matrix B such that for every $A \in \partial f(x_0)$, $\begin{pmatrix} A \\ B \end{pmatrix}$ is invertible?*

This is the problem that has motivated our project.

2 Rademacher's Theorem

In this section, we will present a proof of Rademacher's theorem, following the proof given in [4] and [8]. The reader may wish to skip this section, as the details do not play an important role in the rest of the project. Indeed, the author has included this section so that he may get a better grade for his report! To prove Rademacher's theorem, we will need to define the notion of an absolutely continuous function.

Definition 2.1. *Let I be an interval in the real line $\mathbb{R}$. A function $f : I \to \mathbb{R}$ is* ***absolutely continuous*** *on I if for every $\epsilon > 0$ there exits $\delta > 0$ such that whenever a finite sequence of* ***pairwise disjoint*** *sub intervals (x_k, y_k) of I satisfy*

$$\sum_k |y_k - x_k| < \delta,$$

then

$$\sum_k |f(y_k) - f(x_k)| < \epsilon.$$

Note that a Lipschitz function $f : I \to \mathbb{R}$ is also an absolutely continuous function, a consequence of the the Lipschitz inequality of f.

We have introduced absolutely continuous functions because we will need the following lemma. We will not prove this lemma, but provide a reference for the interested reader.

Lemma 2.2. *Let $f : (a, b) \to \mathbb{R}$ be an absolutely continuous function. Then, f is differentiable almost everywhere with integrable derivative such that*

$$f(b) - f(a) = \int_a^b f'(x)\, dx.$$

Proof. See [7], chapter 7. □

We now state Rademacher's theorem:

Theorem 2.3 (Rademacher's theorem)**.** *Let $f : \mathbb{R}^n \to \mathbb{R}^m$ be Lipschitz. Then, f is differentiable almost everywhere with respect to the Lebesgue measure on $\mathbb{R}^n$ (which we denote by L^n). That is to say, the Jacobian of f, denoted Jf, exists and*

$$\lim_{y \to x} \frac{\|f(y) - f(x) - Jf(x)(y - x)\|}{\|y - x\|} = 0$$

for L^n almost everywhere $x \in \mathbb{R}^n$.

Proof. As a vector valued function is Lipschitz if and only if it is Lipschitz in each component, we may assume without loss of generality that $m = 1$. We denote Jf by ∇f. The proof will be divided into proving three claims.

Fix $v \in \mathbb{R}^n$ with $\|v\| = 1$, and define

$$D_v f(x) := \lim_{t \to 0} \frac{f(x + tv) - f(x)}{t}$$

for $x \in \mathbb{R}^n$ whenever this limit exists.

Claim 1. *$D_v f(x)$ exists for L^n almost everywhere $x \in \mathbb{R}^n$.*

Proof of Claim 1 . As f is Lipschitz, f is continuous, therefore Borel measurable. It follows that

$$\overline{D_v}f(x) := \limsup_{t\to 0} \frac{f(x+tv)-f(x)}{t}$$

is Borel measurable. This may be seen by recalling

- If f_k are Borel measurable for all $k \in \mathbb{N}$, then so is $f(x) := \lim_{k\to\infty} f_k(x)$ if the limit exists almost everywhere,
- If f_k are Borel measurable for all $k \in \mathbb{N}$, then so is $f(x) = \sup_{k\in\mathbb{N}} f_k(x)$.

Similarly,

$$\underline{D_v}f(x) := \liminf_{t\to 0} \frac{f(x+tv)-f(x)}{t}$$

is Borel measurable. Therefore, the set

$$\begin{aligned} B_v &:= \{x \in \mathbb{R}^n \mid D_v f(x) \text{ does not exist}\} \\ &= \{x \in \mathbb{R}^n \mid \underline{D_v}f(x) < \overline{D_v}f(x)\} \end{aligned}$$

is Borel measurable.

Now, for each $x, v \in \mathbb{R}^n$ with $\|v\| = 1$, define $\phi_{x,v} : \mathbb{R} \to \mathbb{R}$ by

$$\phi_{x,v}(t) := f(x+tv).$$

Then, $\phi_{x,v} : \mathbb{R} \to \mathbb{R}$ is Lipschitz, therefore absolutely continuous. Hence, $\phi_{x,v}$ is L^1 differentiable almost everywhere by lemma 2.2. Thus, for each line L parallel to v, for the one-dimensional Lebesgue measure L^1, we have $L^1(B_v \cap L) = 0$. Finally, using Fubini's theorem, for the n dimensional Lebesgue measure L^n, we have

$$L^n(B_v) = \int_{\{\langle x,v\rangle=0\}} L^1(B_v \cap L_x)\, dx = 0$$

where L_x is the line through x parallel to v. As a consequence, we see that

$$\nabla f = (D_1(f), \ldots, D_n f(x))$$

exists for L^n almost everywhere $x \in \mathbb{R}^n$. This concludes the proof of claim 1. □

We now state our second claim:

Claim 2. *We have $D_v f(x) = v \cdot \nabla f(x)$ for L^n almost everywhere $x \in \mathbb{R}^n$.*

Proof of claim 2. Let $C_0^\infty(\mathbb{R}^n)$ denote the set of smooth real valued functions on $\mathbb{R}^n$ with compact support and let $\eta \in C_0^\infty(\mathbb{R}^n)$. Then,

$$\int_{\mathbb{R}^n} \frac{f(x+tv)-f(x)}{t}\eta(x)\, dx = -\int_{\mathbb{R}^n} f(x)\frac{\eta(x)-\eta(x-tv)}{t}\, dx \tag{1}$$

Equation 1 is true because

$$\begin{aligned}
\int_{\mathbb{R}^n} \frac{f(x+tv)-f(x)}{t}\eta(x)\,dx &= \frac{1}{t}\left(\int_{\mathbb{R}^n} f(x+tv)\eta(x)\,dx - \int_{\mathbb{R}^n} f(x)\eta(x)\,dx\right)\\
&= \frac{1}{t}\left(\int_{\mathbb{R}^n} f(y)\eta(y-tv)\,dx - \int_{\mathbb{R}^n} f(x)\eta(x)\,dx\right) \quad (\text{set } y = x+tv)\\
&= \frac{1}{t}\left(\int_{\mathbb{R}^n} f(x)\eta(x-tv)\,dx - \int_{\mathbb{R}^n} f(x)\eta(x)\,dx\right) \quad (y \text{ is a dummy variable})\\
&= -\int_{\mathbb{R}^n} f(x)\frac{\eta(x)-\eta(x-tv)}{t}\,dx.
\end{aligned}$$

Let $t = \frac{1}{k}$ for $k \in \mathbb{N}$ in equation 1 and note

$$\left|\frac{f(x+\frac{1}{k}v)-f(x))}{\frac{1}{k}}\right| \le \mathrm{Lip}(f)\|v\| = \mathrm{Lip}(f) \quad (\text{as } \|v\| = 1). \tag{2}$$

Since f is Lipschitz, f is continuous and hence is bounded on every compact set. As $\eta \in C_0^\infty(\mathbb{R}^n)$, it follows that

$$\int_{\mathbb{R}^n} \frac{f(x+tv)-f(x)}{t}\eta(x)\,dx < \infty.$$

Now, in claim 1, we have shown that $D_v f(x)$ exists for L^n almost everywhere $x \in \mathbb{R}^n$, and have hence established the pointwise convergence for $D_v f(x)$. Furthermore, the sequence of functions $D_v f(x)$ for $t = \frac{1}{k}$ is dominated by $\mathrm{Lip}(f)$ as shown in equation 2. As constants are integrable on compact sets, it follows

$$\int_{\mathbb{R}^n} \mathrm{Lip}(f)|\eta(x)|\,dx < \infty.$$

Thus, by the Dominated Convergence Theorem,

$$\begin{aligned}
\lim_{t\to 0}\int_{\mathbb{R}^n} \frac{f(x+tv)-f(x)}{t}\eta(x)\,dx &= \int_{\mathbb{R}^n} \lim_{t\to 0}\frac{f(x+tv)-f(x)}{t}\eta(x)\,dx\\
&= \int_{\mathbb{R}^n} D_v f(x)\eta(x)\,dx.
\end{aligned}$$

But,

$$\lim_{t\to 0}\int_{\mathbb{R}^n} \frac{f(x+tv)-f(x)}{t}\eta(x)\,dx = \lim_{t\to 0} -\int_{\mathbb{R}^n} \frac{\eta(x)-\eta(x-tv)}{t}f(x)\,dx$$

and $\eta \in C_0^\infty(\mathbb{R}^n)$, so we have

$$\lim_{t\to 0} -\int_{\mathbb{R}^n} \frac{\eta(x)-\eta(x-tv)}{t}f(x)\,dx = -\int_{\mathbb{R}^n} D_v\eta(x)f(x)\,dx.$$

Therefore,

$$\int_{\mathbb{R}^n} D_v f(x)\eta(x)\,dx = -\int_{\mathbb{R}^n} D_v\eta(x)f(x)\,dx.$$

But,

$$\begin{aligned}
-\int_{\mathbb{R}^n} f(x) D_v \eta(x)\, dx &= -\int_{\mathbb{R}^n} f(x) \sum_{i=1}^n v_i \frac{\partial \eta}{\partial x_i}(x)\, dx \\
&= -\sum_{i=1}^n v_i \left(\int_{\mathbb{R}^n} f(x) \frac{\partial \eta}{\partial x_i}(x)\, dx \right).
\end{aligned}$$

Using the fact that an absolutely continuous function is differentiable almost everywhere with integrable derivative (lemma 2.2) and Fubini's theorem, we have

$$\begin{aligned}
\int_{\mathbb{R}^n} f(x) \frac{\partial \eta}{\partial x_i}(x)\, dx &= \int_{\mathbb{R}^{n-1}} \int_{\mathbb{R}} f(x) \frac{\partial \eta}{\partial x_i}(x)\, dx_1 \cdots dx_n \quad \text{(Fubini)} \\
&= \int_{\mathbb{R}^{n-1}} \int_{\mathbb{R}} -\frac{\partial f}{\partial x_i}(x) \eta(x)\, dx_1 \cdots dx_n \quad \text{(Absolute continuity)} \\
&= \int_{\mathbb{R}^n} -\frac{\partial f}{\partial x_i}(x) \eta(x)\, dx \quad \text{(Fubini)}.
\end{aligned}$$

Therefore,

$$\begin{aligned}
\int_{\mathbb{R}^n} D_v f(x) \eta(x)\, dx &= \sum_{i=1}^n v_i \left(\int_{\mathbb{R}^n} \frac{\partial f}{\partial x_i}(x) \eta(x)\, dx \right) \\
&= \int_{\mathbb{R}^n} (v \cdot \nabla f) \eta(x)\, dx.
\end{aligned}$$

As this is true for any $\eta \in C_0^\infty(\mathbb{R}^n)$, it follows that $D_v f(x) = v \cdot \nabla f(x)$ for L^n almost everywhere $x \in \mathbb{R}^n$. This proves claim 2. □

Now, choose a countable dense subset $\{v_k\}_{k=1}^\infty$ of S^{n-1} and define

$$A_k := \{x \in \mathbb{R}^n : \exists\ D_{v_k} f(x), \exists\ \nabla f(x) \text{ and } D_{v_k} f(x) = v_k \cdot \nabla f(x)\}.$$

Also define

$$A := \bigcap_{k=1}^\infty A_k.$$

Note, by claim 2, for any k, $L^n(\mathbb{R}^n \setminus A_k) = 0$. Thus, as A is countable intersection of the A_k, it follows $L^n(\mathbb{R}^n \setminus A) = 0$. We now prove our third claim, which proves Rademacher's theorem:

Claim 3. *f is differentiable at every $x \in A$.*

Proof of claim 3. Define

$$\begin{aligned}
Q : A \times S^{n-1} \times \mathbb{R} \setminus \{0\} &\to \mathbb{R} \\
Q(x, v, t) &:= \frac{f(x + tv) - f(x)}{t} - v \cdot \nabla f(x).
\end{aligned}$$

Then, if $x \in A$ and $v, w \in S^{n-1}$, we obtain

$$\begin{aligned}
|Q(x,v,t) - Q(x,w,t)| &\leq \left| \frac{f(x+tv) - f(x+tw)}{t} \right| + |(v-w) \cdot \nabla f(x)| \quad \text{(triangle inequality)} \\
&\leq \mathrm{Lip}(f)\|v-w\| + \|v-w\|\|\nabla f(x)\| \quad \text{(Lipschitz and Cauchy inequality).}
\end{aligned}$$

But note,

$$\begin{aligned}
|D_{e_j} f(x)| &= \lim_{h \to 0} \left| \frac{f(x+e_j h) - f(x)}{h} \right| \\
&\leq \lim_{h \to 0} \mathrm{Lip}(f) \\
&= \mathrm{Lip}(f).
\end{aligned}$$

Therefore, we obtain the inequality

$$|Q(x,v,t) - Q(x,w,t)| \leq (\sqrt{n}+1)\mathrm{Lip}(f)\|v-w\|. \tag{3}$$

Now, let $\epsilon > 0$. Then, by the compactness of S^{n-1} and the density of $\{v_k\}_{k=1}^{\infty}$, there exists a $N \in \mathbb{N}$ such that for any $v \in S^{n-1}$, there exists a $k < N$ so that

$$\|v - v_k\| \leq \frac{\epsilon}{2(\sqrt{n}+1)\mathrm{Lip}(f)}. \tag{4}$$

Furthermore, by the definition of A, we have $\lim_{t\to 0} Q(x, v_k, t) = 0$. Therefore, there exists a $\delta > 0$ such that

$$|Q(x, v_k, t)| < \frac{\epsilon}{2} \tag{5}$$

for all $0 < |t| < \delta$, for $k < N$. Therefore, for any $x \in S^{n-1}$, there exists a $k < N$ such that

$$\begin{aligned}
|Q(x,v,t)| &\leq |Q(x,v_k,t)| + |Q(x,v,t) - Q(x,v_k,t)| \\
&< \frac{\epsilon}{2} + (\sqrt{n}+1)\mathrm{Lip}(f)\frac{\epsilon}{2(\sqrt{n}+1)\mathrm{Lip}(f)} \quad \text{(by inequalities 3, 4 and 5)} \\
&= \epsilon
\end{aligned}$$

whenever $0 < |t| < \delta$. Also, notice that the same δ works for all $v \in S^{n-1}$. Therefore, for any $x \in A$ and for any $v \in S^{n-1}$, we have

$$\lim_{t \to 0} Q(x,v,t) = 0.$$

Finally, choose any $y \in \mathbb{R}^n$, $y \neq x \in A$ and define

$$v := \frac{y-x}{\|y-x\|}.$$

Then, $y = x + tv$ where $t = \|y - x\|$ and

$$\lim_{y \to x} \frac{|f(y) - f(x) - (y - x) \cdot \nabla f(x)|}{\|y - x\|} = \lim_{y \to x} Q(x, v, \|y - x\|) = 0.$$

This concludes the proof of claim 3, and therefore the proof of Rademacher's theorem. $\square$

$\square$

3 An Inverse and Implicit Function Theorem for Lipschitz Functions using Generalised Derivatives

3.1 Clarke's Inverse Function Theorem

We want to provide an expanded proof of Clarke's Inverse Function Theorem. The ideas of this proof (at least up to the proof of lemma 3.7) will be used later on in the report, so the reader is encouraged to read the proof carefully. We follow the proof given in [2][pp253-255]:

Theorem 3.1. *Let $f : \mathbb{R}^n \to \mathbb{R}^n$ be Lipschitz, $x_0 \in \mathbb{R}^n$. If $\partial f(x_0)$ is of maximal rank, then f is locally invertible with Lipschitz inverse. That is to say, there exists open neighbourhoods U of x_0, V of $f(x_0)$ and Lipschitz $g : V \to \mathbb{R}^n$ such that $g \circ f(u) = u$ for all $u \in U$ and $f \circ g(v) = v$ for all $v \in V$.*

To prove this theorem, we will need the following facts about generalised derivatives. We will not prove them, but will provide a reference for the reader.

Proposition 3.2 (Upper semicontinuity of the generalised derivative)**.** *Let $f : \mathbb{R}^n \to \mathbb{R}^m$ be Lipschitz, $x_0 \in \mathbb{R}^n$. Then, ∂f is upper semicontinuous at x_0. That is to say, for any $\epsilon > 0$ there exits $\delta > 0$ such that for any $y \in \mathbb{B}_\delta(x_0)$,*

$$\partial f(y) \subseteq \Omega_\epsilon.$$

where Ω_ϵ consists of matrices A such that $\|A - B\| < \epsilon$ for at least one $B \in \partial f(x_0)$.

Proof. See Clarke [2][pp 70-71]. $\square$

Proposition 3.3 (Local extrema)**.** *If $g : \mathbb{R}^n \to \mathbb{R}$ is Lipschitz and attains a local minimum or maximum at x_0, then $0 \in \partial g(x_0)$.*

Proof. See Clarke [2][p38]. $\square$

Proposition 3.4 (A chain rule)**.** *Let $h = g \circ f$, where $f : \mathbb{R}^n \to \mathbb{R}^m$ is Lipschitz at x and $g : \mathbb{R}^m \to \mathbb{R}$ is Lipschitz at $f(x)$. Then, h is Lipschitz near x and one has*

$$\partial h(x) \subseteq convH \{\partial g(f(x))\partial f(x)\}$$

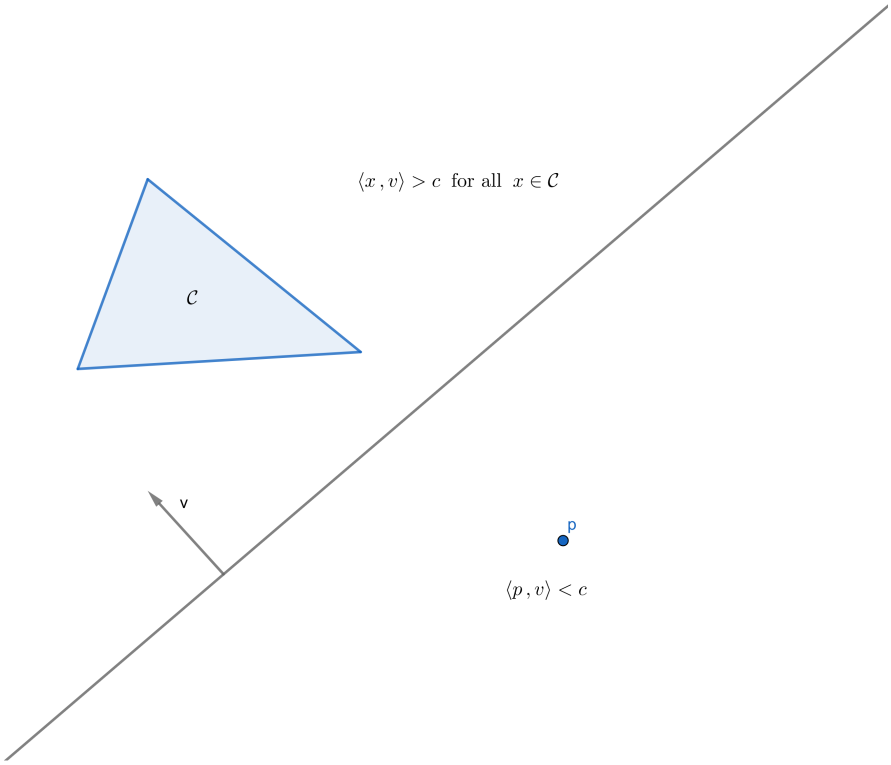

Figure 1: Diagram illustrating the geometry of the separation proposition 3.5.

where the right hand side means the convex hull of all matricies of the form AB, $A \in \partial g(f(x))$, $B \in \partial f(x)$.

Proof. See Clarke [2][pp72-74]. □

In addition, we will also need the following strict separation theorem for a point and a closed convex set:

Proposition 3.5 (Strict separation theorem for a point and a closed convex set)**.** *Let $\mathcal{C}$ be a closed convex subset of $\mathbb{R}^n$ and let $p \in \mathbb{R}^n$ be a point not on $\mathcal{C}$. Then, there exists a nonzero vector v and a real number c such that $\langle x\,,v\rangle > c$ for all $x \in \mathcal{C}$ and $\langle p\,,v\rangle < c$.*

Geometrically, this result says that a closed convex set $\mathcal{C}$ and a point p not on $\mathcal{C}$ may be separated by the affine hyperplane $\{x \in \mathbb{R}^n \mid \langle x\,,v\rangle = c\}$ (see figure 1). We will need this proposition to obtain a lower bound of an inner product when proving the Inverse Function Theorem(see lemma 3.6).

Proof. See [3], example 2.20 [p49]. □

Proof of the Inverse Function Theorem following [2], pp 253-255. The proof will be given in lemma steps.

Lemma 3.6. *Let f and x_0 be as in theorem 3.1. There exists $r, \delta > 0$ such that for any unit vector $v \in \mathbb{R}^n$ there exists a unit vector $w \in \mathbb{R}^n$ such that whenever $x \in \mathbb{B}_r(x_0)$ and $M \in \partial f(x)$, then*

$$\langle w\,, Mv\rangle \geq \delta. \tag{6}$$

Proof of lemma 3.6. Let S^{n-1} be the unit sphere in $\mathbb{R}^n$. Denote by $\partial f(x_0)S^{n-1}$ to mean the union of $A(S^{n-1})$ for every $A \in \partial f(x_0)$. As $\partial f(x_0)$ is of maximal rank, it follows that $\partial f(x_0)S^{n-1} \subseteq \mathbb{R}^n$ does not contain 0. Clarke also states that $\partial f(x_0)S^{n-1}$ is compact. We think this follows from the fact that $\partial f(x_0)$ is of maximal rank or the fact that $\partial f(x_0)$ is compact, but we are not sure. We continue assuming that he is correct. Therefore, there exists a $\delta' > 0$ such that the distance between $\partial f(x_0)S^{n-1}$ and 0 is $2\delta'$. By continuity, for $\epsilon > 0$ sufficiently small, $\Omega_\epsilon S^{n-1}$ (with Ω_ϵ as defined as in proposition 3.2) is distance at least δ' from 0. By proposition 3.2, there exists an $r > 0$ such that for any $x \in \mathbb{B}_r(x_0)$,

$$\partial f(x) \subseteq \Omega_\epsilon. \tag{7}$$

Now, let $v \in \mathbb{R}^n$ be any unit vector. So, $v \in S^{n-1}$. Thus, the set $\Omega_\epsilon v := \{Av \mid A \in \Omega_\epsilon\}$ is a convex set that is distance at least δ' from 0. This is also true for the closure $\overline{\Omega_\epsilon v}$. Thus, by proposition 3.5 applied to $\overline{\Omega_\epsilon v}$ and the origin 0, there exists an unit vector w and a $\delta > 0$ such that

$$\langle w\,, \overline{\Omega_\epsilon v}\rangle \geq \delta.$$

Thus, 6 follows from this and equation 7. □

Lemma 3.7. *Let r, δ be as above. If $x, y \in \overline{\mathbb{B}_r(x_0)}$, then*

$$\|f(x) - f(y)\| \geq \delta\|x - y\|.$$

Proof. Without loss of generality, we may suppose $x \neq y$ and as f is continuous, we may suppose $x, y \in \mathbb{B}_r(x_0)$. Set

$$v := \frac{y - x}{\|y - x\|}$$

and

$$\lambda := \|y - x\|$$

so that

$$y = x + \lambda v.$$

Let Π be the hyperplane perpendicular to v passing through x, see figure 2. By Rademacher's theorem, the set P of points x' in $\mathbb{B}_r(x_0)$ where $Df(x')$ fails to exist is of measure 0. Thus, by Fubini, for almost every x' in Π, the ray,

$$x' + tv$$

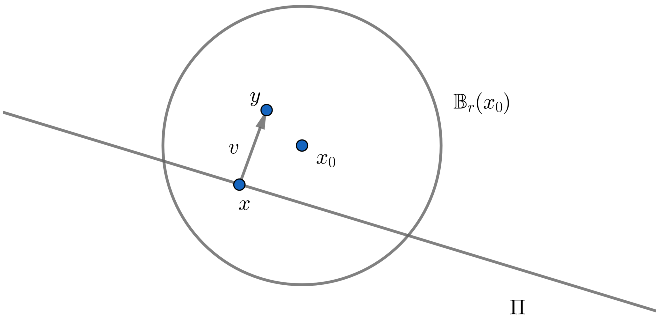

Figure 2: A diagram illustrating what is happening in lemma 3.7.

for $t \geq 0$ meets P in a set of 0 one dimensional measure. Choose x' with the above property and sufficiently close to x such that $x' + tv$ lies in $\mathbb{B}_r(x_0)$ for all $t \in [0, \lambda]$. Therefore, the map

$$\begin{aligned} \phi : [0, \lambda] &\to \mathbb{R}^n \\ t &\longmapsto f(x' + tv) \end{aligned}$$

is Lipschitz and for almost every $t \in [0, \lambda]$, it's classical derivative exists. The classical derivative of this map is $Jf(x'+tv)v$ (this means the jacobian matrix $Jf(x'+tv)$ multiplied by v.) Then, we have

$$f(x' + \lambda v) - f(x') = \int_0^\lambda Jf(x' + tv)v \, dt.$$

A priori, this does not follow immediately from the Fundamental Theorem of Calculus (since the derivative of this function exists only almost everywhere). An example of a function which illustrates the failure of such an equality is the Cantor function. The above equality is true because $\phi : [0, \lambda] \to \mathbb{R}^n$ is Lipschitz and hence Lipschitz in each component. Therefore, it is **absolutely continuous** in each component (see definition 2.1). Thus, by lemma 2.2, each component has integrable derivative such that integrating this derivative on $[0, \lambda]$ gives an equality like in lemma 2.2. The above equality then follows. Let w correspond to v as in Lemma 3.6. We deduce

$$\begin{aligned} w \cdot \big(f(x' + \lambda v) - f(x')\big) &= \int_0^\lambda w \cdot Jf(x' + tv)v \, dt \\ &\geq \int_0^\lambda \delta \, dt \\ &= \delta\lambda \\ &= \delta\|y - x\|. \end{aligned}$$

Thus, as w is an unit vector, it follows that

$$\|f(x' + \lambda v) - f(x'))\| \geq \delta\|x - y\|.$$

This may be done for x' arbitrarily close to x. Since f is continuous, the lemma follows. $\square$

Lemma 3.8. *Let r, δ be as above. Then, $\mathbb{B}_{\frac{r\delta}{2}}(f(x_0)) \subseteq f(\mathbb{B}_r(x_0))$. Moreover, for any $y \in \mathbb{B}_{\frac{r\delta}{2}}(f(x_0))$, there exits an **unique** $x \in \mathbb{B}_r(x_0)$ such that $f(x) = y$.*

Proof. Let $y \in \mathbb{B}_{\frac{r\delta}{2}}(f(x_0))$. Let a local minimum of $\|y - f(\cdot)\|^2$ over $\overline{\mathbb{B}_r(x_0)}$ be attained at x (the minimum of a continuous function on a compact set is always attained). We claim that actually $x \in \mathbb{B}_r(x_0)$. Otherwise, we have

$$\begin{aligned}
\frac{\delta r}{2} &> \|y - f(x_0)\| \quad (\text{as } y \in \mathbb{B}_{\frac{r\delta}{2}}(f(x_0))) \\
&\geq \|f(x) - f(x_0)\| - \|y - f(x)\| \\
&\geq \delta\|x - x_0\| - \|y - f(x)\| \quad (\text{lemma 3.7}) \\
&\geq \delta\|x - x_0\| - \|y - f(x_0)\| \quad (\text{minimality of } x_0) \\
&\geq \delta r - \|y - f(x_0)\| \quad (\text{assumption that } x \notin \mathbb{B}_r(x_0)) \\
&> \delta r - \frac{\delta r}{2} \\
&= \frac{\delta r}{2}.
\end{aligned}$$

This is a contradiction. Thus, $x \in \mathbb{B}_r(x_0)$. If we can show that $y = f(x)$, the inclusion in the lemma is proven.

We have shown that x yields a local minimum for the function $\|y - f(\cdot)\|^2$ within $\mathbb{B}_r(x_0)$. By proposition 3.3, it follows that $0 \in \partial\|y - f(x)\|^2$. We want to apply the chain rule as described in proposition 3.4 to $\partial\|y - f(x)\|^2$. Let

$$\begin{gathered}
F : \mathbb{R}^n \to \mathbb{R}^n \\
F(x) := y - f(x)
\end{gathered}$$

and let

$$\begin{gathered}
g : \mathbb{R}^n \to \mathbb{R} \\
g(z) := z \cdot (y - f(x))\,.
\end{gathered}$$

Then, by proposition 3.4, we have

$$\partial\|y - f(x)\|^2 = \partial(g \circ F)(x) \subseteq \text{convH}\left\{\partial g\left(F(x)\right)\partial F(x)\right\}.$$

We seek to calculate $\partial g(z)$ for any $z \in \mathbb{R}^n$. Note, if we denote $z = (z_1, \ldots, z_n)$, $(y - f(x)) =$

$(y_1 - f_1(x), \ldots, y_n - f_n(x))$, we have

$$\begin{aligned} g(z) &= z \cdot (y - f(x)) \\ &= \sum_{i=1}^{n} z_i(y_i - f_i(x)). \end{aligned}$$

Therefore,

$$\begin{aligned} \frac{\partial g}{\partial z_j} &= \sum_{i=1}^{n} \frac{\partial}{\partial z_j} \left(z_i(y_i - f_i(x)) \right) \\ &= y_j - f_j(x). \end{aligned}$$

Thus, g is classically differentiable everywhere with derivative $Jg(z) = (y - f(x))^t$. This means that for any $z \in \mathbb{R}^n$, we have $\partial g(z) = \{(y - f(x))^t\}$. Therefore,

$$\begin{aligned} \text{convH} \left\{ \partial g \left(F(x) \right) \partial F(x) \right\} &= \text{convH} \left\{ \{(y - f(x))^t\} \partial F(x) \right\} \\ &= (y - f(x))^t \partial F(x) \\ &= (y - f(x))^t \partial (y - f(x)) \\ &= (y - f(x))^t \partial (-f(x)) \\ &= -(y - f(x))^t \partial f(x). \end{aligned}$$

Hence, we conclude that $0 \in -(y - f(x))^t \partial f(x)$. But lemma 3.6 implies that every matrix in $\partial f(x)$ has maximal rank and is therefore invertible. Thus, we conclude that $y = f(x)$. To prove uniqueness, suppose there exists $x_1, x_2 \in \mathbb{B}_r(x_0)$ such that $x_1 \neq x_2$ and $f(x_1) = f(x_2) = y$. Then, $\|f(x_1) - f(x_2)\| = 0$. But this contradicts lemma 3.7. $\square$

Now, to prove the theorem, we define

$$V := \mathbb{B}_{\frac{r\delta}{2}}(f(x_0))$$

and we define the value $g(v)$ for $v \in V$ of function g to be the unique $x \in \mathbb{B}_r(x_0)$ such that $f(x) = v$. We may define

$$U := f^{-1}(V)$$

which is open neighbourhood of x_0 as f is continuous. Finally, $g : V \to U$ is Lipschitz with Lipschitz constant $1/\delta$ because of lemma 3.7. $\square$

3.2 Clarke's Implicit Function Theorem

From this Inverse function theorem, Clarke goes on to prove an Implicit function theorem for Lipschitz functions. To state his Implicit function theorem, we must first give a definition.

Definition 3.9. *Let $f : \mathbb{R}^{n+k} \to \mathbb{R}^k$ be Lipschitz, $(a, b) \in \mathbb{R}^{n+k}$. Use coordinates $(x, y) := (x^1, \ldots, x^n, y^1, \ldots, y^k)$. We define*

$$\Pi_y \partial f(a, b) := \left\{ N \in \mathbb{R}^{k\times k} \mid \textit{there exists } M \in \mathbb{R}^{k\times n} \textit{such that } (MN) \in \partial f(a, b) \right\}.$$

We say $\Pi_y \partial f(a, b)$ is of maximal rank if every matrix $N \in \Pi_y \partial f(a, b)$ has maximal rank.

Corollary 3.10 (Clarke's Implicit function theorem)**.** *Suppose $f : \mathbb{R}^n \times \mathbb{R}^k \to \mathbb{R}^k$ is Lipschitz and $f(a, b) = 0$. Then, if $\Pi_y \partial f(a, b)$ is of maximal rank, then there exists a open neighbourhood $X \subset \mathbb{R}^n$ of a and a Lipschitz function $\xi : X \to \mathbb{R}^k$ such that $\xi(a) = b$ and for any $x \in X$,*

$$f(x, \xi(x)) = 0.$$

Proof following [2] p256. Let $m := n + k$ and consider

$$\begin{aligned} F &: \mathbb{R}^m \to \mathbb{R}^m \\ F(x, y) &= (x, f(x, y)) . \end{aligned}$$

When the Jacobian matrix JF exists, it is of the form $\begin{pmatrix} I & 0 \\ J_x f & J_y f \end{pmatrix}$. It follows that $\partial F(a, b)$ is of maximal rank. Thus, by Clarke's Inverse function theorem, there exits open neighbourhoods U of (a, b); V of $F(a, b) = (a, f(a, b)) = (a, 0)$ and a Lipschitz $G : V \to U$ such that $F \circ G = id_V$ and $G \circ F = id_U$. Shrinking U and V if necessary, we may assume the V may be written as the product of two open neighbourhoods $V = X \times Y$, $X \subseteq \mathbb{R}^n$, $Y \subseteq \mathbb{R}^k$.

Suppose $G(x, y) = (G_1(x, y), G_2(x, y))$. Then, we have

$$\begin{aligned} (x, y) &= F \circ G(x, y) \\ &= F (G_1(x, y), G_2(x, y)) \\ &= (G_1(x, y), f (G_1(x, y), G_2(x, y))) . \end{aligned}$$

Therefore, $x = G_1(x, y)$ and $y = f(x, G_2(x, y))$. Setting $y = 0$ we deduce that

$$0 = f(x, G_2(x, 0))$$

and setting $x = a$, we deduce that

$$G_2(a, 0) = b.$$

Thus, if we define

$$\begin{aligned} \xi &: X \to \mathbb{R}^k \\ \xi(x) &:= G_2(x, 0), \end{aligned}$$

we have $\xi(a) = b$ and $f(x, \xi(x)) = 0$. $\square$

4 Another Implicit Function Theorem using Generalised Derivatives

We now prove another Implicit function theorem, which is similar to Clarke's Implicit function theorem but has a slightly weaker hypothesis and conclusion. We will present two proofs of the theorem: one longer direct proof using ideas taken from Lee's Introduction to smooth manifolds [5][pp661-662], and the other using Clarke's Implicit function theorem.

Theorem 4.1. *Let $U \subseteq \mathbb{R}^{n+k}$ be open; $(x, y) = (x^1, \cdots, x^n, y^1, \cdots, y^k)$; $f : U \to \mathbb{R}^k$ Lipschitz at (a, b) in U and define $c := f(a, b)$. Then, if there exists a $n \times (n + k)$ matrix B such that for every $A \in \partial f((a, b))$, $\begin{pmatrix} A \\ B \end{pmatrix}$ is invertible, then $f^{-1}(c)$ is locally a Lipschitz submanifold of dim n.*

Proof 1 using ideas from Lee's introduction to smooth manifolds. As $\begin{pmatrix} A \\ B \end{pmatrix}$ is invertible for every $A \in \partial f((a, b))$, it follows that B is of maximal rank. Thus, using elementary row and column operations, we may deduce that there exists invertible matrices P, Q such that

$$PBQ = \begin{pmatrix} I_n & 0_{n\times k} \end{pmatrix}.$$

Define $\tilde{B} := \begin{pmatrix} I_n & 0_{n\times k} \end{pmatrix}$ and define $\tilde{f} := f \circ Q$. Thus $\tilde{f} : Q^{-1}(U) \to \mathbb{R}^k$. Let $(\tilde{a}, \tilde{b}) := Q^{-1}((a, b))$. We claim that

$$\partial \tilde{f}((\tilde{a}, \tilde{b})) = \{\tilde{A} := AQ \mid A \in \partial f((a, b))\}.$$

Indeed,

$$\begin{aligned} \partial \tilde{f}((\tilde{a}, \tilde{b})) &= \operatorname{convH}\left(\left\{\lim_{m\to\infty} d_{Q^{-1}(x_m,y_m)}(f \circ Q) \mid (x_m, y_m) \to (a, b)\right\}\right) \\ &= \operatorname{convH}\left(\left\{\lim_{m\to\infty} d_{(x_m,y_m)} f \circ Q \mid (x_m, y_m) \to (a, b)\right\}\right) \\ &= \{AQ \mid A \in \partial f((a, b))\}. \end{aligned}$$

Now, we define

$$\begin{aligned} \tilde{\psi} &: Q^{-1}(U) \to \mathbb{R}^{n+k} \\ \tilde{\psi}(x, y) &:= \left(P^{-1} \circ \tilde{B}(x, y), \tilde{f}(x, y)\right) \\ &= \left(P^{-1}(x), \tilde{f}(x, y)\right). \end{aligned}$$

Then, one may show that $\partial\tilde{\psi}((\tilde{a},\tilde{b}))$ consists of invertible matrices

$$\left\{ \begin{pmatrix} B \\ A \end{pmatrix} Q \mid A \in \partial f((a,b)) \right\}.$$

Hence, by Clarke's Inverse Function Theorem, there exists open neighbourhoods $\tilde{U}_0$ of $(\tilde{a},\tilde{b})$ and $\tilde{Y}_0$ of $\tilde{\psi}((\tilde{a},\tilde{b})) = (P^{-1}(\tilde{a}), c)$ such that $\tilde{\psi} : \tilde{U}_0 \to \tilde{Y}_0$ is invertible with Lipschitz inverse. Shrinking if necessary, we may assume $\tilde{U}_0$ is the product of neighbourhoods $\tilde{U}_0 = \tilde{V} \times \tilde{W}$.

Now, suppose

$$\tilde{\psi}^{-1}(x,y) = \left(\tilde{C}(x,y), \tilde{D}(x,y)\right)$$

for some functions $\tilde{C} : \tilde{Y}_0 \to \tilde{V}$, $\tilde{D} : \tilde{Y}_0 \to \tilde{W}$. Then,

$$\begin{aligned}
(x,y) &= \tilde{\psi}\left(\tilde{\psi}^{-1}(x,y)\right) \\
&= \tilde{\psi}\left(\tilde{C}(x,y), \tilde{D}(x,y)\right) \\
&= \left(P^{-1}\left(\tilde{C}(x,y)\right), \tilde{f}\left(\tilde{C}(x,y), \tilde{D}(x,y)\right)\right).
\end{aligned}$$

Thus, $P^{-1}\left(\tilde{C}(x,y)\right) = x$, and so $\tilde{C}(x,y) = P(x)$. Therefore,

$$\tilde{\psi}^{-1}(x,y) = \left(P(x), \tilde{D}(x,y)\right).$$

We define $\tilde{V}_0 := \left\{x \in \tilde{V} \mid \left(P^{-1}(x), c\right) \in \tilde{Y}_0\right\}$ and let $\tilde{W}_0 = \tilde{W}$ (note, $\tilde{V}_0$ is an open neighbourhood of $(P^{-1}(\tilde{a}), c)$). Further, let us define

$$\begin{gathered}
\tilde{F} : \tilde{V}_0 \to \tilde{W}_0 \\
\tilde{F}(x) = \tilde{D}(P^{-1}(x), c).
\end{gathered}$$

We have

$$\begin{aligned}
\left(P^{-1}(x), c\right) &= \tilde{\psi} \circ \tilde{\psi}^{-1}\left(P^{-1}(x), c\right) \\
&= \tilde{\psi}\left(x, \tilde{D}\left(P^{-1}(x), c\right)\right) \\
&= \tilde{\psi}\left(x, \tilde{F}(x)\right) \\
&= \left(P^{-1}(x), \tilde{f}(x, \tilde{F}(x))\right).
\end{aligned}$$

Therefore, $c = \tilde{f}(x, \tilde{F}(x))$ when $x \in \tilde{V}_0$. Hence, graph $\tilde{F} \subseteq \tilde{f}^{-1}(c)$.

Now, suppose $(x, y) \in \tilde{V}_0 \times \tilde{W}_0$ such that $\tilde{f}(x, y) = c$. We have

$$\begin{aligned}\tilde{\psi}(x, y) &= \left(P^{-1}(x), \tilde{f}(x, y)\right) \\ &= (P^{-1}(x), c).\end{aligned}$$

Thus,

$$\begin{aligned}(x, y) &= \tilde{\psi}^{-1}\left(P^{-1}(x), c\right) \\ &= \left(x, \tilde{D}(P^{-1}(x), c)\right) \\ &= \left(x, \tilde{F}(x)\right).\end{aligned}$$

Therefore, $\tilde{f}^{-1}(c)\,|_{\tilde{V}_0 \times \tilde{W}_0} \subseteq \text{graph } \tilde{F}$. Thus, we have showed that $\tilde{f}^{-1}(c)$ is the graph of some Lipschitz function on some open domain. Therefore, $\tilde{f}^{-1}(c)$ is locally a Lipschitz submanifold.

However,

$$\begin{aligned}\tilde{f}^{-1}(c) &= (f \circ Q)^{-1}(c) \\ &= Q^{-1}\left(f^{-1}(c)\right).\end{aligned}$$

Therefore, $Q^{-1}\left(f^{-1}(c)\right)$ is locally a Lipschitz submanifold. But Q is an isomorphism. Thus

$$Q\left(Q^{-1}\left(f^{-1}(c)\right)\right) = f^{-1}(c)$$

is also locally a Lipschitz submanifold. □

Proof 2 using Clarke's Implicit function theorem. In the notation of the above proof, $\partial\tilde{\psi}(\tilde{a}, \tilde{b})$ consists of invertible matrices

$$\begin{aligned}&\left\{\begin{pmatrix} B \\ A \end{pmatrix} Q \mid A \in \partial f((a, b))\right\} \\ &= \left\{\begin{pmatrix} P^{-1} & 0_{n\times k} \\ \multicolumn{2}{c}{AQ} \end{pmatrix} \mid A \in \partial f((a, b))\right\}.\end{aligned}$$

Therefore, the far right $k \times k$ minor of AQ must have maximal rank. That is to say, $\Pi_y \partial\tilde{f}(\tilde{a}, \tilde{b})$ is of maximal rank. Thus, by Clarke's Implicit function theorem, there exists an open neighbourhood $\tilde{X} \subseteq \mathbb{R}^n$ of $\tilde{a}$ and a Lipschitz function $\tilde{\xi} : \tilde{X} \to \mathbb{R}^k$ such that $\tilde{\xi}(\tilde{a}) = \tilde{b}$ and $\tilde{f}(\tilde{x}, \tilde{\xi}(\tilde{x})) = c$ for all $\tilde{x} \in \tilde{X}$. Thus, we have showed that $\tilde{f}^{-1}(c)$ is locally the graph of some Lipschitz function i.e. it is locally a Lipschitz submanifold. However,

$$\begin{aligned}\tilde{f}^{-1}(c) &= (f \circ Q)^{-1}(c) \\ &= Q^{-1}\left(f^{-1}(c)\right).\end{aligned}$$

Therefore, $Q^{-1}\left(f^{-1}(c)\right)$ is locally a Lipschitz submanifold. But Q is an isomorphism. Thus

$$Q\left(Q^{-1}\left(f^{-1}(c)\right)\right) = f^{-1}(c)$$

is also locally a Lipschitz submanifold. $\square$

Note, we **cannot** apply Clarke's Implicit function theorem directly, as A being of maximal rank k does not imply that its far right $k \times k$ minor is of maximal rank. For example, consider the 2×3 matrix $\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \end{pmatrix}$.

5 Clarke's Implicit Function Theorem implies Wuertz's Implicit Function Theorem

5.1 Wuertz's Implicit Function Theorem

The Implicit function theorem given by Clarke is not the only Implicit function theorem given for Lipschitz functions. Michael Wuertz in his Masters Thesis states and proves a different Implicit function theorem for Lipschitz functions. The theorems of Clarke and Wuertz give the same conclusion but have different hypotheses. It seems as if Wuertz was unaware of Clarke's theorem, as he does not reference him within his bibliography.

Wuertz's Implicit function theorem (see [8]) is as follows:

Theorem 5.1 (Wuertz's Implicit function theorem)**.** *Let $U_n \subseteq \mathbb{R}^n$ and $U_k \subseteq \mathbb{R}^k$ be open. Next, fix $a \in U_n$ and $b \in U_k$ and define $U := U_n \times U_k$. Consider*

$$F : U_n \times U_k \to \mathbb{R}^k$$

a Lipschitz function such that

$$F(a, b) = 0$$

and with the property that there exists a constant $K > 0$ for which

$$\|F(x, y_1) - F(x, y_2)\| \geq K\|y_1 - y_2\| \quad \textit{for all} \quad (x, y_j) \in U.$$

Then, there exists $V_n \subseteq \mathbb{R}^n$ open, such that $a \in V_n$, and a Lipschitz function $\varphi : V_n \to U_k$ such that $\varphi(a) = b$, and

$$\{(x, y) \in V_n \times U_k : F(x, y) = 0\} = \{(x, \varphi(x)) : x \in V_n\}.$$

In particular,

$$F(x, \varphi(x)) = 0, \quad \textit{for all} \quad x \in V_n.$$

Therefore, the difference between Clarke's Implicit function theorem and Wuertz's Implicit function theorem is in their respective hypotheses. Specifically, Clarke requires

$\Pi_y f(a,b)$ to be of maximal rank, whereas Wuertz requires $\|F(x,y_1)-F(x,y_2)\| \geq K\|y_1 - y_2\|$ for all $(x,y_j) \in U$. A priori, there appears to be no relation between the two conditions. However, examining the proof of Clarke's Inverse function theorem carefully, we see that he proves a similar inequality in his second lemma.

We have adapted his arguments to prove that if $\Pi_y f(a,b)$ is of maximal rank, then we have the inequality $\|F(x,y_1) - F(x,y_2)\| \geq K\|y_1 - y_2\|$ for all $(x,y_j) \in U$, where U is an open neighbourhood of (a,b).

5.2 A Proof of the Implication

The proof of this implication is very similar to the proof of lemma 3.7 in Clarke's Inverse Function theorem. The reader is encouraged to read the proof of 3.7, as he/she will better understand the proof given here having read that proof.

Proposition 5.2. *Let $f : \mathbb{R}^{n+k} \to \mathbb{R}^k$ be Lipschitz, $(a,b) \in \mathbb{R}^{n+k}$ such that $f(a,b) = 0$ and $\Pi_y \partial f(a,b)$ has maximal rank. Then, there exists an open neighbourhood U of (a,b) and $K > 0$ such that*

$$\|f(x,y_1) - f(x,y_2)\| \geq K\|y_1 - y_2\|$$

for all $(x,y_j) \in U$.

Proof. The proof will consist of two lemma steps, similar to the steps as in the proof of theorem 3.1.

Lemma 5.3. *There exists $r, K > 0$ such that for any unit vector $(v_1, \ldots, v_{n+k}) \in \mathbb{R}^{n+k}$ such that $(v_1, \ldots, v_n) = 0$, there exists unit vector $w \in \mathbb{R}^k$ such that whenever $(x,y) \in \mathbb{B}_r(a,b)$ and $M \in \partial f(a,b)$,*

$$\langle w, Mv \rangle \geq K.$$

Proof of lemma 5.3 . Let S^{n+k-1} be the unit sphere in $\mathbb{R}^{n+k}$.

Define $\tilde{S}^{n+k-1} := \{v \in S^{n+k-1} \mid (v_1, \ldots, v_n) = 0\}$. As $\Pi_y \partial f(a,b)$ is of maximal rank, it follows $(\Pi_y \partial f(a,b))\, \tilde{S}^{n+k-1}$ does not contain 0 and is compact. Hence, there exists a $K' > 0$ such that the distance between $(\Pi_y \partial f(a,b))\, \tilde{S}^{n+k-1}$ and 0 is $2K'$.

Define Ω_ϵ to be the set containing $(n+k) \times k$ matrices η such that $\|\eta - \xi\| < \epsilon$ for at least one $\xi \in \partial f(a,b)$.

By continuity, for $\epsilon > 0$ sufficiently small, $\Omega_\epsilon \tilde{S}^{n+k-1}$ is distance at least K' from 0. By the upper semicontinuity of the generalised derivative (see proposition 3.2), there exists $r > 0$ such that for any $(x,y) \in \mathbb{B}_r(a,b)$,

$$\partial f(x,y) \subseteq \Omega_\epsilon. \tag{8}$$

Now, let $v \in \tilde{S}^{n+k-1}$. Then, $\Omega_\epsilon v$ is a convex set that is distance at least K' from 0. This is also true for $\overline{\Omega_\epsilon v}$. By proposition 3.5 applied to $\overline{\Omega_\epsilon v}$ and 0, there exists an unit vector

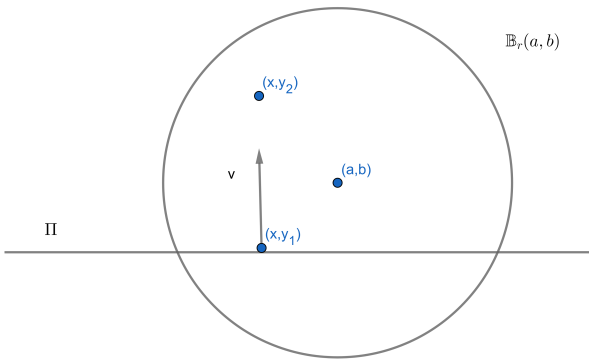

Figure 3: A diagram illustrating what is happening in lemma 5.4].

$w \in \mathbb{R}^k$ and $K > 0$ such that

$$\langle w, \overline{\Omega_\epsilon v} \rangle \geq K.$$

The lemma follows from this and equation 8. ☐

Lemma 5.4. *Let* $(x, y_1), (x, y_2) \in \mathbb{B}_r(a, b)$. *Then,*

$$\|f(x, y_1) - f(x, y_2)\| \geq K\|y_1 - y_2\|.$$

Proof of lemma 5.4. Without loss of generality, $y_1 \neq y_2$. Let

$$v := \frac{(0, y_1 - y_2)}{\|y_1 - y_2\|}$$

and

$$\lambda := \|y_1 - y_2\|.$$

Hence, $(x, y_1) = (x, y_2) + \lambda v$.

Let Π be the hyperplane perpendicular to v passing through (x, y_2) (see figure 3). The set P of points (x', y') in $\mathbb{B}_r(a, b)$ where $Df(x', y')$ fails to exist is of measure 0, hence by Fubini's theorem, for almost everywhere $(x', y') \in \Pi$, the ray

$$(x', y') + tv, \quad t \geq 0$$

meets P in a set of 0 one dimensional measure. Choose a (x', y') with the above property and sufficiently close to (x, y_2) so that $(x', y') + tv$ lies in $\mathbb{B}_r(a, b)$ for all $t \in [0, \lambda]$. Then, the function

$$t \mapsto f((x', y') + tv)$$

is Lipschitz and differentiable almost everywhere. It's derivative is $Jf((x',y')+tv)v$. Then, as explained in the proof of lemma 3.7 we have

$$f((x',y')+\lambda v)-f((x',y'))=\int_0^\lambda Jf((x',y')+tv)v\,dt.$$

Let $w\in\mathbb{R}^k$ be as in lemma 5.3. Then,

$$\begin{aligned}w\cdot\big(f((x',y')+\lambda v)-f((x',y'))\big)&=\int_0^\lambda w\cdot(Jf((x',y')+tv)v)\,dt\\&\geq\int_0^\lambda K\,dt\\&=K\|y_1-y_2\|.\end{aligned}$$

Thus, as w is an unit vector, it follows that

$$\|f((x',y')+\lambda v)-f((x',y'))\|\geq K\|y_1-y_2\|.$$

By continuity, we have

$$\|f(x,y_1)-f(x,y_2)\|\geq K\|y_1-y_2\|.$$

□

This concludes the proof of the proposition, thus showing that if the hypothesis of Clarke's Implicit function theorem are satisfied, then the hypothesis of Wuertz's Implicit function theorem are satisfied. □

5.3 A Partial Converse

It is natural to ask whether Wuertz's hypothesis implies Clarke's hypothesis. We believe that this should be true, as a priori we see no reason why Wuertz's Implicit function theorem should be superior to Clarke's Implicit function theorem. However, we were only able to prove this implication in the special case where f is classically differentiable at (a,b) and $\Pi_y\partial f(a,b)=\{\partial_y f(a,b)\}$, where $\partial_y f(a,b)$ is the far right $k\times k$ minor of $d_{(a,b)}f$. Maybe our method could be generalised to prove the general case, or maybe a completely different approach is required.

We prove the contrapositive:

Proposition 5.5. *Let $f:\mathbb{R}^{n+k}\to\mathbb{R}^k$ be Lipschitz and classically differentiable at $(a,b)\in\mathbb{R}^{n+k}$. Suppose $\Pi_y\partial f(a,b)=\{\partial_y f(a,b)\}$, where $\partial_y f(a,b)$ is the far right $k\times k$ minor of $d_{(a,b)}f$.*

*If $\partial_y f(a,b)$ is **not** invertible, then for any open neighbourhood U of (a,b) and for any*

$K > 0$, *there exists* $(x, y_1), (x, y_2) \in U$ *such that*

$$\|f(x, y_1) - f(x, y_2)\| < K\|y_1 - y_2\|.$$

Proof. It suffices to prove that there exists $(a, y_i)_{i=1}^{\infty}$ such that

$$(a, y_i) \to (a, b)$$

and

$$\frac{\|f(a, y_{i+1}) - f(a, y_i)\|}{\|y_{i+1} - y_i\|} \to 0.$$

As $\partial_y f(a, b)$ is not invertible, it follows that there exists a **nonzero** $v \in \mathbb{R}^k$ such that $\partial_y f(a, b)v = 0$. Let us consider the vector $(0, v) \in \mathbb{R}^{n+k}$. Then, by the definition of the classical derivative,

$$\lim_{\lambda \to 0 \text{ in } \mathbb{R}} \frac{\|f(a, b + \lambda v) - f(a, b) - \lambda d_{(a,b)} f(0, v)\|}{\|\lambda v\|} = 0.$$

But, by construction, $d_{(a,b)} f(0, v) = 0$. Therefore, we have

$$\lim_{\lambda \to 0 \text{ in } \mathbb{R}} \frac{\|f(a, b + \lambda v) - f(a, b)\|}{\|\lambda v\|} = 0.$$

Define the sequence $(a, y_i) \subseteq \mathbb{R}^{n+k}$ where

$$y_i := \begin{cases} b & i \text{ odd} \\ b + \frac{1}{i} v & i \text{ even.} \end{cases}$$

Then, we have $(a, y_i) \to (a, b)$ and, by the above, we have

$$\frac{\|f(a, y_{i+1}) - f(a, y_i)\|}{\|y_{i+1} - y_i\|} \to 0.$$

□

6 The Motivating Problem

We now want to consider the following question: Let $U \subseteq \mathbb{R}^n$ be open; $n > p$; $f : U \to \mathbb{R}^p$ Lipschitz and $x_0 \in U$. Then, if for every $A \in \partial f(x_0)$, there exists a $p \times n$ matrix B_A such that $\begin{pmatrix} A \\ B_A \end{pmatrix}$ is invertible, is it true that there exists a $p \times n$ matrix B such that for every $A \in \partial f(x_0)$, $\begin{pmatrix} A \\ B \end{pmatrix}$ is invertible? Recall that we are considering this problem as it offers a criterion to apply the Implicit function theorem stated before in this report. We did not manage to prove the result or give a counterexample, but we were able to prove a weaker result.

6.1 A Weak Result about The Motivating Problem

We were able to prove that if for every $A \in \partial f(x_0)$ there exists a $(n-p) \times n$ matrix B_A such that $\begin{pmatrix} A \\ B_A \end{pmatrix}$ is invertible, then there exists a **bounded** finite set of $(n-p)\times n$ matrices $\{B_1, \dots, B_m\}$ such that for every $A \in \partial f(x_0)$, $\begin{pmatrix} A \\ B_i \end{pmatrix}$ is invertible for some $1 \le i \le m$. Our proof will show that if $p = 1$, then $m = 1$ and if $p > 1$, then $m \le \prod_{k=1}^{p-1}(n-k)$. The ideas required to prove this are no more sophisticated than compactness, but we can conclude for example that for $n = 3$ and $p = 2$, we only require a set of two matrices $\{B_1, B_2\}$ so that for every $A \in \partial f(x_0)$, either $\begin{pmatrix} A \\ B_1 \end{pmatrix}$ is invertible or $\begin{pmatrix} A \\ B_2 \end{pmatrix}$ is invertible.

The proof is given as an inductive argument, for which the next proposition acts as the base case:

Proposition 6.1. *Let $\mathcal{C} \subseteq \mathbb{R}^n$ be compact and convex such that $0 \notin \mathcal{C}$. Then, there exists a $(n-1)\times n$ matrix B such that for every $v \in \mathcal{C}$, $\begin{pmatrix} v^t \\ B \end{pmatrix}$ is invertible.*

Proof. Let $\mathcal{C} \subseteq \mathbb{R}^n$ be compact and convex such that $0 \notin \mathcal{C}$. We claim that there exists an unique $\hat{v} \in \mathcal{C}$ which is closest to the origin (see figure 4). Indeed, let us define

$$\begin{aligned} f &: \mathcal{C} \to \mathbb{R} \\ f(v) &:= \|v\|. \end{aligned}$$

Then, f is a continuous function on a compact set. Therefore, f achieves a minimum. This proves the existence of $\hat{v}$. To prove uniqueness, suppose v_1 and v_2 are minima of f. Consider $(1-t)v_1 + tv_2$ for $t \in [0, 1]$. This is in $\mathcal{C}$ for all t as $\mathcal{C}$ is convex. Thus, $0.5v_1 + 0.5v_2$ is closer to the origin than v_1 or v_2, which is of course a contradiction unless $v_1 = v_2$.

Now consider the hyperplane $\mathcal{H}$ through the origin that is orthogonal to $\hat{v}$. Suppose $\mathcal{H}$ has basis $\{b_1, \dots, b_{n-1}\}$. The claim is that

$$B := \begin{pmatrix} \cdots b_1^t \cdots \\ \vdots \\ \cdots b_{n-1}^t \cdots \end{pmatrix}$$

works. Indeed, it suffices to prove that no $v \in \mathcal{C}$ lies in the span of $\{b_1, \dots, b_{n-1}\}$.

Suppose otherwise. Let $v \in \mathcal{C}$ lie in the span of $\{b_1, \dots, b_{n-1}\}$. Then, as $\mathcal{C}$ is convex, it follows that $(1-t)v + t\hat{v} \in \mathcal{C}$ for all $t \in [0, 1]$. Therefore, $0.5v + 0.5\hat{v}$ is closer to the origin than $\hat{v}$. This is a contradiction.

□

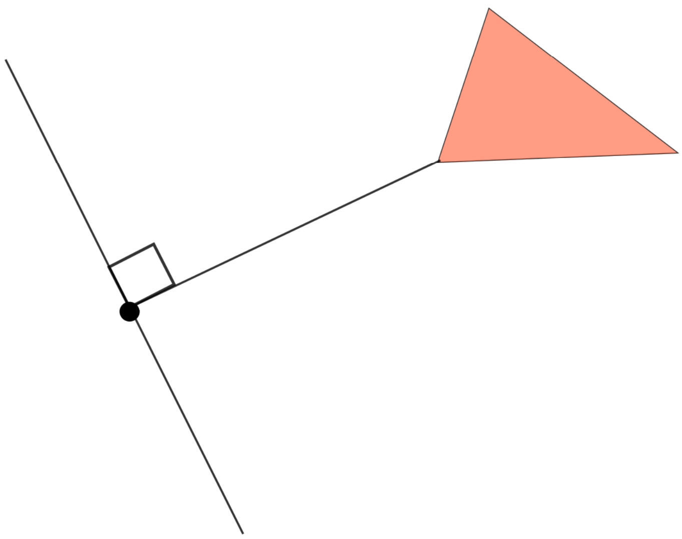

Figure 4: Showing the hyperplane through the origin orthogonal to the vector with smallest size.

We have proven the base case of the inductive proof of the following proposition:

Proposition 6.2. *Suppose $U \subseteq \mathbb{R}^n$ is open, $f : U \to \mathbb{R}^p$ is Lipschitz at $x_0 \in U$ and $n > p$. If for every $A \in \partial f(x_0)$ there exists a $(n-p) \times n$ matrix B_A such that $\begin{pmatrix} A \\ B_A \end{pmatrix}$ is invertible, then there exists a* ***bounded*** *finite set of $(n-p) \times n$ matrices $\{B_1, \ldots, B_m\}$ such that for every $A \in \partial f(x_0)$, $\begin{pmatrix} A \\ B_i \end{pmatrix}$ is invertible for some $1 \leq i \leq m$. Our proof will show that for $p > 1$, $m \leq \prod_{k=1}^{p-1}(n-k)$.*

Proof. We prove by induction on p. If $p = 1$, the result follows from proposition 6.1. Thus, in this case, we may take $m = 1$.

Suppose we have proven the result for everything less than p. We want to prove the result for p. Let $\pi_{\mathbb{R}^{p-1}} : \mathbb{R}^p \to \mathbb{R}^{p-1}$ be the projection onto the first $p-1$ coordinates. Define $\hat{f} := \pi_{\mathbb{R}^{p-1}} \circ f$. Thus, $\hat{f} : U \to \mathbb{R}^{p-1}$. We claim the following: every $\hat{A} \in \partial \hat{f}(x_0)$ may be realised as an $A \in \delta_{x_0} f$ with its last row deleted. Recall,

$$\partial \hat{f}(x_0) = \text{convH}\left(\left\{\lim_{m\to\infty} d_{x_m}\hat{f} \mid x_m \to x_0\right\}\right).$$

It suffices to prove every $\hat{M} = \lim_{m\to\infty} d_{x_m}\hat{f}$ may be realised as an $M = \lim_{i\to\infty} d_{x_i} f$ with its last row deleted. Let

$$\begin{aligned}
\hat{M} &= \lim_{m\to\infty} d_{x_m}\hat{f} \\
&= \lim_{m\to\infty} d_{x_m}\left(\pi_{\mathbb{R}^{p-1}} \circ f\right) \\
&= \lim_{m\to\infty} \left(\pi_{\mathbb{R}^{p-1}} \circ d_{x_m} f\right).
\end{aligned}$$

It is not necessarily true that $\lim_{m\to\infty} d_{x_m} f$ exists. However, by sequential compactness, there exists a subsequence $(x_{m_k})_{k=1}^{\infty}$ such that $\lim_{k\to\infty} d_{x_{m_k}} f$ exists. Note, $\hat{M} = \lim_{k\to\infty} d_{x_{m_k}} \hat{f}$. Thus, defining $M := \lim_{k\to\infty} d_{x_{m_k}} f$, it follows $\pi_{\mathbb{R}^{p-1}} \circ M = \hat{M}$.

Furthermore, by similar reasoning as above, it follows that deleting the last row of some $A \in \partial f(x_0)$ gives some $\hat{A} \in \partial \hat{f}(x_0)$.

Now, from the above, it follows that every $\hat{A} \in \partial \hat{f}(x_0)$ is of maximal rank, as its corresponding $A \in \partial f(x_0)$ is of maximal rank by hypothesis. Therefore, there exists a $(n-p+1) \times n$ matrix $B_{\hat{A}}$ such that $\begin{pmatrix} \hat{A} \\ B_{\hat{A}} \end{pmatrix}$ is invertible. Thus, by the inductive hypothesis, there exists a finite set of $(n-p+1) \times n$ matrices $\{B_1, \ldots, B_r\}$ such that for every $\hat{A} \in \partial \hat{f}(x_0)$, $\begin{pmatrix} \hat{A} \\ B_i \end{pmatrix}$ is invertible for some $B_i \in \{B_1, \ldots, B_r\}$.

Now, let $A \in \partial f(x_0)$ with

$$A = \begin{pmatrix} \cdots a_1^t \cdots \\ \vdots \\ \cdots a_p^t \cdots \end{pmatrix}.$$

Then,

$$\hat{A} = \begin{pmatrix} \cdots a_1^t \cdots \\ \vdots \\ \cdots a_{p-1}^t \cdots \end{pmatrix}.$$

Let

$$B_i = \begin{pmatrix} \cdots b_1^t \cdots \\ \vdots \\ \cdots b_{n-p+1}^t \cdots \end{pmatrix}$$

be the matrix such that $\begin{pmatrix} \hat{A} \\ B_i \end{pmatrix}$ is invertible. Then, as $a_1, \ldots, a_{p-1}, b_1, \ldots, b_{n-p+1}$ form a basis of $\mathbb{R}^n$ and $a_1, \ldots, a_p$ are linearly independent, it follows there exists constants such that

$$a_p = \lambda_1 a_1 + \cdots + \lambda_{p-1} a_{p-1} + \mu_1 b_1 + \cdots + \mu_{n-p+1} b_{n-p+1}$$

where $\mu_j \neq 0$ for some j. Thus,$\{a_1, \ldots, a_p, b_1, \ldots, b_{j-1}, b_{j+1}, \ldots, b_{n-p+1}\}$ are linearly independent. Thus,

$$\begin{pmatrix} A \\ B_{i,j} \end{pmatrix}$$

is invertible where

$$B_{i,j} := \begin{pmatrix} \cdots b_1^t \cdots \\ \vdots \\ \cdots b_{j-1}^t \cdots \\ \cdots b_{j+1}^t \cdots \\ \vdots \\ \cdots b_{n-p+1}^t \cdots \end{pmatrix}.$$

$\square$

6.2 Difficulties Trying to Solve The Motivating Problem

Having discovered a positive result for the motivating problem in the special case $p = 1$ (see proposition 6.1), we attempted to generalise our arguments in the hope of finding a positive result for general p.

Specifically, for $n > p$, consider $g : U \subseteq \mathbb{R}^n \to \mathbb{R}^p$ Lipschitz and let $x_0 \in U$. Define $\mathcal{C} := \partial g(x_0)$ and define

$$\Sigma := \{M \in \mathbb{R}^{p\times n} \mid \text{rank}(M) < p\}.$$

We considered the function

$$d : \mathcal{C} \times \Sigma \to \mathbb{R}$$
$$d(A, \sigma) := \|A - \sigma\|^2$$

where the norm is understood to be the Euclidean distance norm. This function is intended to generalise the function f considered in proposition 6.1. It was relatively straightforward to prove that f in proposition 6.1 achieved an unique minimum. However, this property is not clear for the function d.

Being optimistic and assuming that d does achieve an unique minimum in all cases, we then tried to mimic our proof of proposition 6.1. Specifically, if we denote the unique minimum of d by $(\hat{A}, \hat{\sigma})$, we considered the **orthogonal complement** of the minimal length vector $v := \hat{A} - \hat{\sigma}$ within the space $\mathbb{R}^{p\times n}$ (the reader should perhaps pause here and consider why this mimics the proof of proposition 6.1). This vector subspace of $\mathbb{R}^{p\times n}$, denoted $\mathcal{H}$, has dimension $pn - 1$.

In the proof of proposition 6.1, the dimension of $\mathcal{H}$ was $n-1$ and a basis set of $\mathcal{H}$ was sufficent to construct our desired $n - 1 \times n$ matrix B. However, in the general case, the reader will see that a basis set of $\mathcal{H}$ consists of $pn - 1$ **matrices**. It is not clear how one constructs a $n - p \times n$ matrix B from a set of $pn - 1$ matrices, each of dimension $p \times n$.

Maybe a completely new method is required, or maybe we were not smart enough to make such a construction!

7 Conclusion

To summarise, this dissertation was motivated by the following problem: Let $U \subseteq \mathbb{R}^n$ be open; $n > p$; $f : U \to \mathbb{R}^p$ Lipschitz and $x_0 \in U$. Then, if for every $A \in \partial f(x_0)$, there exists a $p \times n$ matrix B_A such that $\begin{pmatrix} A \\ B_A \end{pmatrix}$ is invertible, is it true that there exists a $p \times n$ matrix B such that for every $A \in \partial f(x_0)$, $\begin{pmatrix} A \\ B \end{pmatrix}$ is invertible? Although we could not find a proof or a suitable counterexample to this problem, we did manage to prove a weaker result. We proved that, under the same hypothesis, that there exists a **bounded** finite set of $(n-p) \times n$ matrices $\{B_1, \ldots, B_m\}$ with $m = 1$ if $p = 1$ and $m \leq \prod_{k=1}^{p-1}(n-k)$ if $p > 1$ such that for every $A \in \partial f(x_0)$, $\begin{pmatrix} A \\ B_i \end{pmatrix}$ is invertible for some $1 \leq i \leq m$. If the author had more time, he would try to find a counterexample to the original problem, as he conjectures that the statement of the problem is false.

In addition, we managed to prove a weaker Implicit function theorem for Lipschitz functions stated in terms of generalised derivatives. We provided two proofs: one using the ideas from Lee's Introduction to Smooth Manifolds [5] and another proof using Clarke's Implicit function theorem 3.10.

However, in the author's opinion, the most interesting part of the dissertation was studying the link between Clarke's Implicit function theorem 3.10 and Wuertz's Implicit function theorem 5.1. Specifically, we proved that if for Lipschitz $f : \mathbb{R}^{n+k} \to \mathbb{R}^k$,

$$\Pi_y \partial f(a,b) := \left\{ N \in \mathbb{R}^{k \times k} \mid \text{there exists } M \in \mathbb{R}^{k \times n} \text{such that } (MN) \in \partial f(a,b) \right\}$$

consists of invertible $k \times k$ matrices, then, there exists an open neighbourhood U of (a,b) and $K > 0$ such that

$$\|f(x,y_1) - f(x,y_2)\| \geq K\|y_1 - y_2\|$$

for all $(x, y_j) \in U$. Naturally, we asked whether the converse was true, but we only managed to prove the converse in the special case where f is classically differentiable at (a,b) and $\Pi_y \partial f(a,b)$ only contains the far right $k \times k$ minor of the Jacobian $Jf(a,b)$. The author conjectures that the converse implication should be true, as he sees no reason why Wuertz's Implicit function theorem 5.1 should be superior to Clarke's Implicit function theorem 3.10. If the author had more time, he would definitely try to prove this result.

Thus, the original problem remains unsolved and we have discovered another interesting question!

References

[1] Francis H. Clarke. On the inverse function theorem. *Pacific Journal of Mathematics*, 1976.

[2] Francis H. Clarke. *Optimization and Nonsmooth Analysis.* Society for Industrial and Applied Mathematics, 1990.

[3] Stephen Boyd et al. *Convex Optimization.* Cambridge University Press, 2004.

[4] Xiaoping Yang Fanghua Lin. *Geometric Measure Theory: An Introduction.* International Press, Boston, 2002.

[5] John M Lee. *Introduction to Smooth Manifolds.* Springer, second edition, 2012.

[6] R.T.Rockafellar. *Convex Analysis.* Princeton University Press, 1970.

[7] Walter Rudin. *Real and Complex Analysis.* McGraw HiII Book Company, third edition, 1987.

[8] Michael Wuertz. The implicit function theorem for lipschitz functions and applications. Master's thesis, University of Missouri, 2008.